

Unit 4: Correlation

Marius Ionescu

09/20/2011

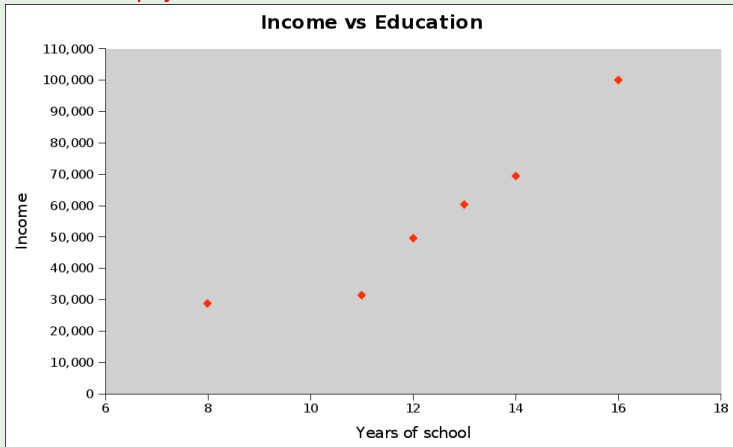
Fact

Correlation = the relationship between two variables

Scatter Diagram

Example

Education pays:

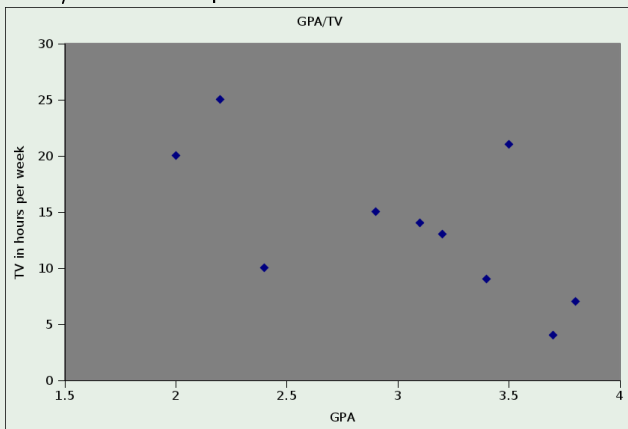


We see a **positive** correlation

Another scatter diagram

Example

GPA/TV hourse per week:



We see a negative correlation.

Scaling a scatter diagram

Fact

- *The best way to choose a scale for a scatter diagram is to use standard units*
- *If x and y_i is your data point and the average of x is $\bar{x}$ and the average of y is $\bar{y}$ with SD_x and SD_y then convert to standard units as follows*

$$X = \frac{x_i - \bar{x}}{SD_x} \text{ and } Y = \frac{y_i - \bar{y}}{SD_y}.$$

How do we measure the correlation?

Definition

- The **correlation coefficient** is

$$r = \text{average}(XY).$$

- That is, the correlation coefficient equals the average of (X in standard units times Y in standard units).

Fact

- *The correlation coefficient is always between -1 and 1 .*
- *We standardize it so that ± 1 is perfectly linear, and 0 is void of any linear relation.*

Example

Example

Find the correlation of

x	y
1	4
2	5
6	4
11	15

Exceptional Cases

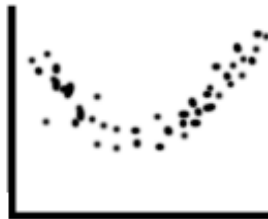
Fact

The r -value can be misleading when we have

- *outliers*
- *nonlinear association*



CLEAR ASSOCIATION
BUT $R = 0$



NONLINEAR ASSOC.
BUT $R = 0$

Fact

r is valid when the scatterplot is football-shaped!

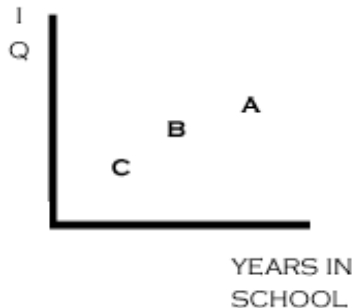
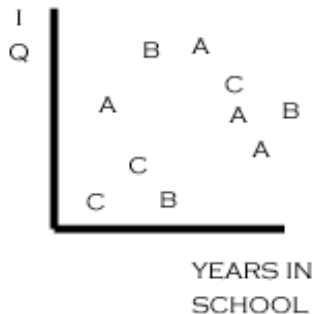
Fact

- *Are based on rates (e.g., averages).*
- *Used often in Poly Sci and Sociology.*
- *They tend to overstate the strength of an association.*
- *Reason: When we average we get rid of variations.*

Example

Example

Consider IQ versus years in school, separated by age group: A = 30 year olds; B = 40 year olds; C = 50 year olds



Association is not Causation

Fact

Important:

“Correlation does not imply causation”

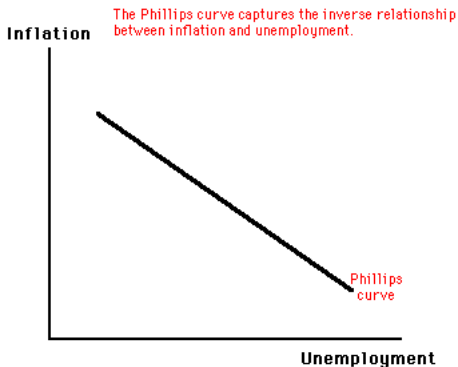
Example

Are people rich because they go to school or they go to school because they are rich?

Example

Example (Phillips Curve)

- Phillips curve is a historical inverse relationship between the rate of unemployment and the rate of inflation in an economy.
- Stated simply, the lower the unemployment in an economy, the higher the rate of inflation.



Example (Phillips Curve)

- The data fit a line very well from 1940 to late 1960's.
- It was used in policy and we got high inflation and high unemployment.

Example

Example

Studies showed that there is a strong correlation between crime rate and temperature in Chicago.

If the mayor puts 200 more police officers in the street will the temperature go down?